Xinmian Huang

Brian Tong

Ashley Yu

CS591 - Data Mechanics

Improving Student Safety

Introduction

      Boston has multiple K-12 schools, colleges, and universities, many of which are ranked top in the nation. Even so, as Boston is a major metropolis, it is also filled with crimes, from minor infractions to serious felonies. As we walk to class along Commonwealth Avenue, we can often hear the sirens of police cars and emergency vehicles racing down the streets. Our phones are filled with warning messages from BUPD about crime incidences nearby. If we do have a crime to report, we can either go to the BUPD which is nearby or use one of the emergency phone lines to contact them.

      However, not everywhere in Boston has such easy access to the police and so our goal is to find the optimal locations where we should build student help centers. These help centers are there so that students who are passing through the area and felt threatened (being followed, harassed, etc) can come to the help centers if they didn't feel comfortable going into the police station or were unable to call the police. To decide on the locations of the help centers, we wanted to analyze the concentration of these crimes in relation to police stations in the area. We hoped to combine these data sets so we can see if there is any correlation between the prevalence of crime and distance to the police stations in Boston. Our hypothesis was that there would be a higher concentration of crime near police stations. But, we also do not want to build students help centers too close to these stations as it would be useless having a help station right beside an actual police station.

Datasets for Project 2:

- Boston Police Stations:

  http://bostonopendata-boston.opendata.arcgis.com/datasets/e5a0066d38ac4e2abbc7918197a4f6af_6

  This dataset contains the location of the police stations within the City of Boston

- Crimes:

  https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

  This dataset contains the crimes committed within the City of Boston. It contains data such as the the type of crime, the location, date and time of the crime.

- Public Schools:

  https://data.boston.gov/dataset/public-schools

  This dataset contains the locations of public schools.

- Non-Public Schools

 https://data.boston.gov/dataset/non-public-schools

  This dataset contains the locations of non-public schools

Data Transformations

First, we found the optimal locations for student help centers using the k-means algorithm. We believe help centers should be located in areas where crime rates are higher, and should be at least one km away from any police departments (otherwise it would be useless). In addition, no more than four help centers are allowed considering budgets and rental spaces. We used k-means algorithm to find geographical centers of areas with higher frequency of crime. Then we filtered out centers that doesn't satisfy our constraints. The remaining centers are help center locations. Below is the visualization of where the help centers (red) would be located in relation to police stations (blue). We chose no more than four help centers for this project, but if budget

and necessity allows, we could alter the algorithm to provide more than four locations. However, we believe only four are necessary as it provides students help without overcrowding or competing with police stations.
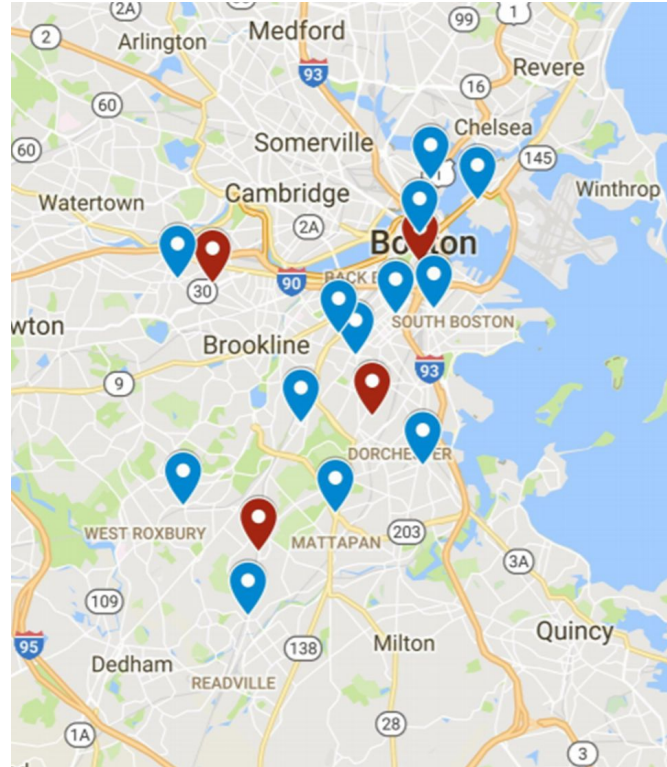
Next, we wanted to find out if the presence of police stations reduces crime in nearby areas. Our hypothesis was that there is more crime near police stations due to the fact that the police stations may have been built because of the frequency of crime in the area. If it is the case that police stations and crime are strongly correlated, then it gives us more reason to build help centers away from police stations in order to improve student safety.

We calculated the correlation coefficient between distance to closest police department and number of crimes in the area. Distances are categorized in ranges of 0.5 km (i.e. 5 ~ 5.5 km). We found a correlation of around -0.865 which shows that the closer to a police station, the higher crime rate in the area, supporting our hypothesis This does not necessarily imply that police "attracts" crimes. A probable explanation is that police departments are usually in areas of higher population, where crime naturally

occurs more often. However, now we know that the presence of police stations doesn't help reduce crime. Therefore, we should build help centers a suitable distance away from police stations in order to help students feel safe when reporting their crimes.
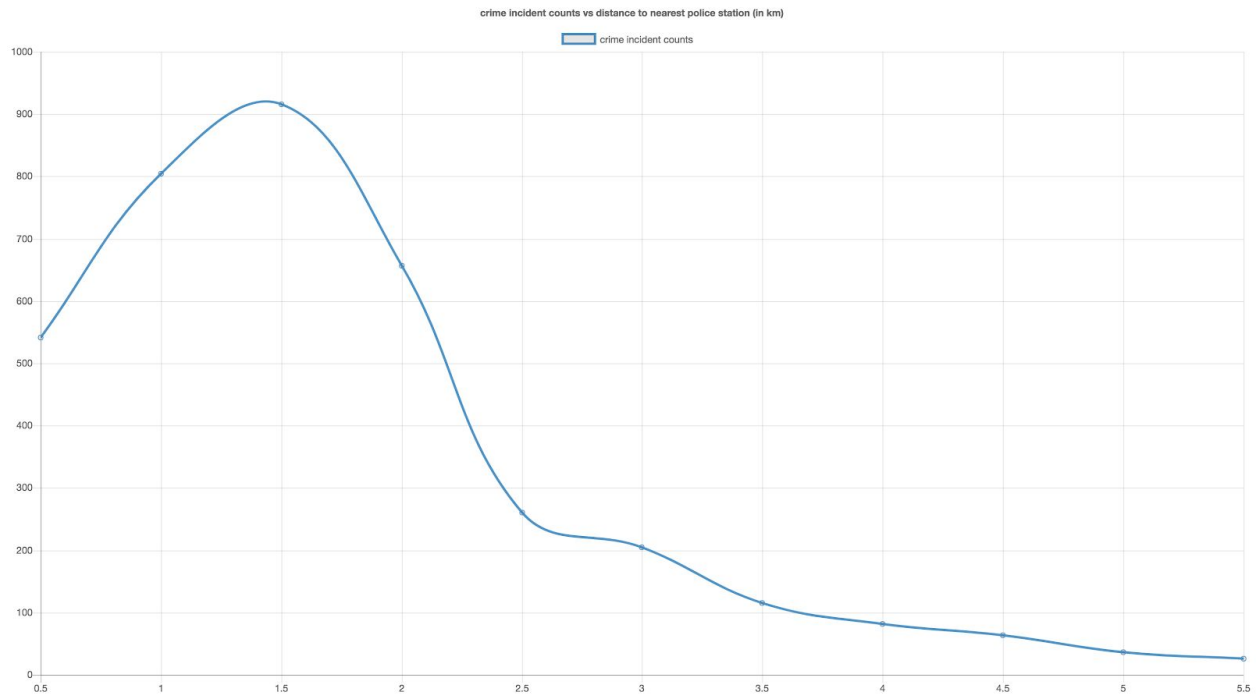
Visualization

Our interactive web-based visualization are composed of two parts and three graphs. To use this RESTful API, run webservice.py, and you should be able to use the service on localhost:5000. To see the visualization, use the /client endpoint, in other words localhost:5000/client. If you're required to login, use 'xhug' for both username and password.

The first part is a google map service that shows the locations of our future student help centers. You can interact with it just like you would with a normal google map: zoom in, drag, click on locations, etc. The only exception is to search for other locations, which I find irrelevant to our project.

The second part is two graphs to visualize the frequency of crime relative to its distance to the nearest police station. The first graph is a pie chart, and the second graph is a line chart, which was shown above. The two styles of visualization helps give the viewer a direct perception of where the majority of crime take place, and how the distance to police affect crime rate.

Extensions

Moving forward with this project, we would like to research more factors to help students feel safer in the city. Ideally, we would have liked to also include property values in the project so we can compare the property values in relation to crime. If we are able to figure out where the "safe" areas are in a city, we can produce information for students, especially students coming from out of town, so that they are aware of where they can go for help, and what places they should avoid if they can.

We would have also liked to include information about the locations of schools around the area, mainly colleges and universities as they would have the highest amount of non-local students. With this information we can determine where are the best locations to place an optimum amount of student help centers so that they are closer to schools for easier access and in safer areas.

Of course, this data can change within a few years as neighborhoods become gentrified. According to an article by Real Estate Boston.com, from 2000-2015, 21.1% of census tracts that were eligible to gentrify did so. In response, real estate prices rose over 50% within these 15 years. This process can greatly change the effect of student help centers as neighborhoods change in value and therefore the concentration of crime changes to. The full article can be found here:

http://realestate.boston.com/news/2016/02/02/boston-gentrification-over-the-past-15-years-in-one-map/

Conclusion

For a basic implementation we feel that this project produced results worthy of looking further into. With the extensions mentioned above we could provide a safer environment for students in the Boston area and improve the city's livelihood.